心理科学进展 2022, Vol. 30, No. 4, 834–850
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2022.00834

预期错误在恐惧记忆更新中的作用与机制*

李俊娇[1] 陈 伟[2,3,4] 石 佩[2,3,4] 董媛媛[2,3,4] 郑希付[2,3,4]

([1]广东第二师范学院教师教育学院, 广州 510303)
([2]华南师范大学心理学院; [3]华南师范大学心理应用研究中心, 广州 510631)
([4]广东省心理健康与认知科学重点实验室, 广州 510631)

摘 要 依据错误驱动的学习理论, 行为预期结果与实际结果之间的不匹配即预期错误(Prediction error, PE)是学习产生的驱动因素。作为显著性信息中的一种, 预期错误和物理显著性、惊讶、新异性等存在信息加工阶段的不同, 与记忆更新的关系也有差异。近年来, 记忆再巩固干预范式(reconsolidation interference)被证明可用于人类条件性恐惧记忆的更新, 其中记忆提取激活阶段所包含的预期错误起到了引发记忆“去稳定”、开启记忆再巩固的关键作用。在促进恐惧记忆更新的行为机制上, PE 被认为是记忆去稳定的必要非充分条件。记忆提取必须包含适量的 PE, 但其引发的是记忆去稳定、消退还是中间状态, 还需结合记忆本身性质确定。在促进恐惧记忆更新的神经机制上, 杏仁核、导水管周围灰质(PAG)、海马均在 PE 探测和计算过程中具有重要作用; 前额叶皮层(PFC)及其亚区在 PE 开启记忆再巩固过程中扮演了重要角色。上述过程又受到神经系统中特定神经递质的重要调节, 尤其是多巴胺能和谷氨酸能。未来研究应进一步探索基于 PE 计算模型的量化研究, 整合 PE 与其他边界条件的交互作用, 考察不同类型显著性在记忆再巩固中的作用等; 并亟待使用多学科手段探索 PE 在恐惧记忆更新中作用的神经与分子机制。同时, 需进一步开展 PE 作用的个体差异研究, 促进研究结果向临床应用转化。

关键词 预期错误, 条件性恐惧, 记忆更新, 再巩固, 提取干预范式
分类号 B845

恐惧症(Phobia)、焦虑障碍(Anxiety)和创伤后应激障碍(Post-traumatic stress disorders, PTSD)是我国罹患精神疾患人群的主要类型, 近年频发的各类自然灾害、突发性公共卫生事件等, 容易给危机事件当事人和目击者造成不同程度的心理障碍。以消退训练为原理的暴露治疗是目前恐惧症和焦虑障碍临床治疗的主要方法之一, 但其存在复发率较高的问题。研究表明传统的消退训练没有消除或更新原有记忆, 而是建立起一种新的安全记忆, 与原始恐惧记忆相竞争, 因而可能在多种情况下出现复发。近年来出现的以记忆再巩固(Reconsolidation)理论为基础的恐惧记忆提取消退(Retrieval-Extinction)范式, 被证明可以有效消除恐惧记忆并抑制复发, 其作用在于通过引发并干扰记忆去稳定(Destabilization)状态, 阻止其再次巩固(Re-stabilization), 从而达到破坏原始记忆联结的目的。但该范式的应用受限于再巩固的边界条件, 即存在一些条件限制了记忆有效激活进入不稳定状态, 其中记忆提取阶段至关重要的边界条件是预期错误(Prediction Error, PE)。因此, 对于预期错误在恐惧记忆更新(包括消除和改写)中的作用及其神经机制的研究, 对于深化此类理论难题的解决和促进实验室结果向临床治疗转化都显得尤为重要。

收稿日期: 2021-06-25
* 国家自然科学基金项目(32000752, 31970996), 广东省哲学社会科学规划项目(GD19YXL01), 广东省教育科学“十三五”规划项目(2019JKDY025), 教育部人文社会科学研究项目(20YJC190009)。
通信作者: 郑希付, E-mail: zhengxifu@m.scnu.edu.cn

1 错误驱动的学习理论与预期错误计算模型

1.1 错误驱动的学习(Error-driven learning)理论

错误驱动的学习理论认为, 刺激物所带来的

强化物必须是让人感到惊讶或不可预测的，个体才会形成学习。当某种行为引起了意料之外的结果，个体就会产生新的学习；如果实际结果完全符合预期，与脑中存储的原始记忆相吻合，个体就不会产生新学习；而如果已经习得的行为不再带有预期的结果，则该行为就会消退(Schultz, 2000)。预期结果与实际结果之间的差异或不匹配(mismatch)，被称为预期错误(PE)。因此，学习本质上是一个由“错误”驱动的过程。

体现这一学习原理的包括条件反射作用中著名的阻碍(blocking)现象。例如，当一个声音多次以匹配食物的方式出现时，仅呈现声音也会引起动物的唾液分泌，表明已经形成了声音−食物联结。而如果此时在出现声音的时候再配合灯光，即声音和灯光同时出现来跟随食物；在形成“声+光”−食物的联结之后，仅仅呈现灯光刺激，却不会引起唾液分泌，说明并未形成灯光−食物联结。因为基于先前的学习，声音已经能够单独预测食物，在后续学习中凭借这一经验也足以预测当下的强化物，和预期完全相符，因而其他与之相匹配的刺激出现并不能引起新的学习，新联结或修改原有联结都不会发生(Schultz et al., 1997)。

在恐惧消退学习中，当原本引起负性结果的条件刺激(conditioned stimulus, CS)不再预测负性结果时，个体会逐步形成一种 CS 匹配安全信息的记忆，体现出个体面对不断变化的环境的一种适应性。可见消退学习或安全学习中的一个重要因素就是预期的变化。当预期和结果非常一致时，个体原先的经验可以充分发挥作用，新的学习就不会发生。只有当预期和结果出现不一致时，个体发现原先的经验对当下的环境已不再起作用时，才会出现学习的动力。可见在该理论下，PE 对恐惧习得和消退等过程都有重要作用，是学习产生的基本驱动。

1.2　预期错误的理论模型

预期错误的经典计算模型主要包括三类，分别是 Rescorla & Wagner 模型、Pearce-Hall 模型，以及时间预期错误(temporal difference, TD)的模型。这三类模型既是预期错误模型，又是关于学习的模型(model of learning)。

Rescorla-Wagner 理论(简称 RW 模型)认为，在条件化学习中，学习由实际的无条件刺激(unconditioned stimulus, US)强度与预期 US 强度之间的错误信号来控制(Rescorla & Wagner, 1972)。将实际 US 列为 λ，预期的 US 列为 ΣV (表示所有 CS 跟随 US 的联结强度 V 的总和)，则错误信号即可列为 λ−ΣV。那么在这一条件下反映在联结强度中的突触修改和变化的学习规则就可以用以下方程表示，S 为学习率：

$$\Delta V = S(\lambda - \Sigma V)$$

Pearce-Hall 理论(简称 PH 模型)认为，学习只会在强化物是令人惊讶(surprising)的时候发生(Pearce & Hall, 1980)。这一模型使用的是奖赏性预期错误(reward prediction error, RPE)的绝对值，其计算方法与未标记的 RPE (unsigned RPE，指 PE 不区分正性和负性)一致。PH 模型认为，错误信号会调节个体分配给每个条件化中试次的注意的多少。如果将注意大小列为 α，它在这个 CS 试次(n)上的量是与上一个试次中预期错误的大小成比例的，可以通过下面的方程来表示：

$$\alpha_n = |\lambda - \Sigma V|_{n-1}$$

这一模型下突触可塑性修改或联结强度变化的信号则可以表示为：

$$\Delta V = \alpha_n S \lambda$$

如果在上一个试次中 CS 对 US 的预期效力很低，则 α 就会较大，那么突触可塑性修改的动力就强；反之 α 就小，则当前试次中联结强度变化的动力就低(McNally et al., 2011)。

时间性预期错误模型(TD)认为，当对在两个时间点(t, t + 1)上出现奖赏的预期变化时，学习也会发生。TD 模型的计算原理与前两种模型不同，它并非计算实际 US 和预期 US 之间的错误信号，而是将预期 US 强度和实际 US 强度加起来(时间 t 上的 $\lambda_t + \Sigma V_t$)，进而比较这一总和的正面价值对比前一时刻上预期的 US 强度，因此 TD 中的错误信号可以表示为以下方程：

$$\Delta V = S\left(\lambda_t + \Sigma V_t - \Sigma V_{t-1}\right)$$

其中 t 和 t−1 是两个连续的时刻，因此 TD 被定义为实际的或预期的 US 强度或者二者之和，超过了之前这个时间点上所预期的 US 强度(Ergo et al., 2020; McNally et al., 2011)。

不难看出，这三类学习模型各有侧重，PH 模型不区分预期错误的方向，只考虑绝对差异；RW 模型将实际强化物与预期差异的方向考虑在内，考虑大于预期或小于预期的情况，区分 PE 的类型，同时强调外部强化的作用；而 TD 模型则重视

内部强化与强化出现的时间点, 认为时间信息可以传递预期错误, 扩展了 PE 的范围和形式。这三种理论尤其是 RW 模型和 TD 模型对于预期错误驱动的学习与记忆研究产生了深远的影响。

1.3 预期错误的类型

总结以往研究可以发现, 在不同的学习模型下 PE 有不同的分类, 主要有以下几种:

第一, RPE 和惩罚性预期错误(punishment prediction error, PPE)。在操作性条件反射(operational conditioning)学习中, PE 根据行为的结果效价(valence)分为两种类型: 奖赏预期错误和惩罚预期错误。Schultz 指出, 奖赏(reward)是一种操作性概念, 用以描述赋予一个物体、一个行为动作或一种内在物理状态的积极属性(Schultz et al., 1997)。与刺激相关联的奖赏价值是一种非刺激本身属性, 不是天然具有的属性。这两类 PE 中, RPE 信号与中脑多巴胺(dopamine, DA)系统密切相关, 而 PPE 信号通路则被认为与精神障碍相关。尤其是 RPE 与中脑多巴胺能神经元(dopaminergic neuron)的关系, 得到了大量研究的验证(Colombo, 2014; Kim et al., 2014; Papalini et al., 2020; Starkweather et al., 2017)。

第二, 正性预期错误(positive PE)和负性预期错误(negative PE)。在以巴普洛夫经典条件(Pavlovian conditioning)反射为基础的条件性恐惧学习中, 根据实际出现 US 与预期 US 的相对大小, 可以分为正性 PE 和负性 PE 两种。前者通常指实际出现的 US 比预期的更大而产生的 PE, 后者指实际出现的 US 比预期的小而产生的 PE, 包括 US 缺失(absent)。负性 PE 可以通过增大预期或者降低 US 强度来达成, 其中最简单的一种模式就是消退, 表现在消退过程中, CS 单独出现却没有跟随 US。

有研究者认为, 正性 PE 导致恐惧学习, 负性 PE 导致恐惧消退。Rescorla 等人认为, 实际产生的强化强度小于原始强化物的情况下, 也会产生负性 PE (Rescorla & Wagner, 1972)。为了提升恐惧消退效果, 一个可能的策略是使 PE 最大化。负性 PE 越大, 学习的动力越强; 随着预期的逐步调整, PE 逐步减小而趋近于 0 时, 学习也会逐渐完成直到停止, 达成某种习得。

第三, 有标记的预期错误(singed prediction error, SPE)和未标记的预期错误(unsigned prediction error, UPE)。按照有无标记(sign), 可以将预期错误分为两类。标记通常指数字前的符号(+、−), 用在 PE 中表示预期错误的效价, 对应正性或者负性(Ergo et al., 2020)。如果明确表示出实际出现的结果比预期的更大或者更小, 则属于有标记的 PE (SPE); 如果仅仅表示实际结果与预期的不同、不匹配, 而没有明确方向的话, 则属于未标记的 PE (UPE)。研究发现, 这两类预期错误具有不同的神经基础和加工机制。按照 PE 的三种计算模型, RW 模型和 TD 模型中的学习规则都是依赖于 SPE。如果 SPE 的方向为正, 则多巴胺释放增多, 若方向为负, 则多巴胺消耗减少。

对比奖赏性 PE 和正性 PE、惩罚性 PE 和负性 PE, 区别在于两者从属的记忆模型不同, 含义也有差别。奖赏性 PE 和惩罚性 PE 多见于工具性条件反射模型(instrumental conditioning), 根据行为引发的结果属性来区分。而正性 PE 和负性 PE 多见于经典条件反射模型(Pavlovian conditioning), 缺乏操作性行为, 根据 CS-US 的关系进行界定。预期 CS 会跟随 US, 结果却没有出现; 或实际出现的 US 比预期的小, 属于负性 PE。而预期 CS 不会跟随 US, 结果却出现 US; 或实际出现的 US 比预期的大, 则属于正性 PE。因而在理论上存在奖赏性的负性 PE 或惩罚性的正性 PE 的可能性。最后, 有标记的 PE 和未标记的 PE 则是在不区分具体学习模型、仅就有无方向进行的分类。

1.4 预期错误与其他类型显著性的关系

出于进化的原因, 个体的注意会被显著的或凸显的刺激所吸引, 这种情况统称为显著性(salience)。显著性是一个宏观概念, 从本质上来说, 任何类型的 PE 都具有显著性。研究者认为, 不同类型的显著性体现了对刺激的不同加工深度(Diederen & Fletcher, 2021)。由于显著性的概念涵盖了多种类型, 包括物理显著性、惊讶、新异性、不完整线索等, 近年来出现在相关文献中容易引起混乱, 因此有必要将这些概念及作用加以澄清。以从属关系上来看, 显著性包括了新异性、效价评估、稀有性和其他显著性。其中新异性又包括物理显著性、惊讶(即意外的新异性)和非意外的新异性等。惊讶对应了未标记的预期错误(UPE), 效价加工对应标记的预期错误(SPE) (图 1)。

这些不同类型的显著性之间遵循两种逻辑关系: 第一, 从属于信息加工的不同阶段。个体对

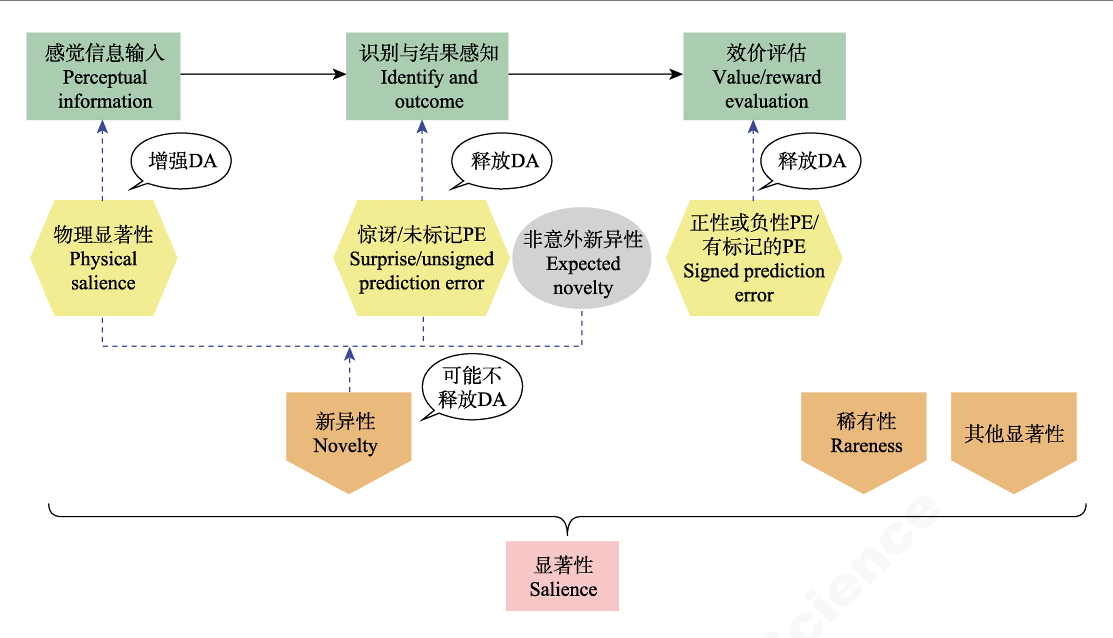


图 1　显著性相关概念关系示意图

注：显著性包括了刺激的新异性、效价评估、刺激稀有性和其他显著性。其中新异性又包括物理显著性、惊讶(即意外的新异性)和非意外的新异性，而只有意外的新异性才会引发多巴胺的释放。如果仅有物理显著性而与结果无直接关系的话，仅能引起多巴胺短暂增强，不足以引起释放。而预期错误(包括 UPE 和 SPE)则主要是涉及了识别与结果感知或效价评估过程。

刺激的加工，经历了感觉信息输入、对刺激结果的初步知觉、对刺激引发或伴随的强化物的效价加工与评定等环节，这三个方面都可能存在显著性信息；第二，对应着中脑多巴胺短暂增强或大量释放的不同情况(图 1)。

1.4.1　物理显著性(physical salience)

感觉输入上存在显著性的物理刺激被证明可以引起极为迅速的多巴胺神经元的相位兴奋(phasic)，响应速度可达 50~110 ms，例如 VTA 中多巴胺能神经元对光刺激的响应。由于这一短暂时间不足以引起细节的加工、识别与评估，因而一般认为物理显著性与作为结果的奖赏或强化物没有直接关联，尽管物理显著的刺激存在得到强化的可能性。也有人认为，物理刺激显著性本身就是一种强化，因为存在进化上的优势，可以让个体尽快识别可能的危险物等(Diederen & Fletcher, 2021)。

1.4.2　惊讶(surprise)

一般认为，惊讶是包含结果属性的，而不仅仅是像物理显著性一样单纯只是感觉输入信息。但是惊讶不包含效价，仅仅表示实际出现结果与预期存在差异，也就是说这一结果属性不带有正性或负性的方向，因此是一种未标记的预期错误(UPE)。有研究证明，惊讶在脑中的加工区域和有标记的预期错误(SPE)存在不同；前者在脑中主要激活前额叶上部，而后者主要加工区域在纹状体或中脑区域。因而在一些研究中会将惊讶与 PE 等同，考察其引发记忆不稳定状态的效果(Sinclair & Barense, 2018)。

1.4.3　新异性(novelty)

物理显著性和惊讶都属于新异性，而新异性又是显著性中的一种；与新异性相比，显著性是一个更广泛的概念。通常在面临新异刺激时，多巴胺神经元会增加，而一旦新异性刺激变得熟悉了并且没有得到强化，则多巴胺释放就会因习惯化而减少。人类功能性磁共振(functional magnetic resonance imaging, fMRI)研究证明，脑中的黑质(substantia nigra, SN)/VTA 区域对新异性刺激有反应，而其他类型的显著性例如稀有性(rareness)和负性情绪等，都不会引起此区域的反应。对于新异性的加工包括早期识别和后期加工，两个过程受到多巴胺的影响不同。最后，新异性并非在所有情况下都会引起多巴胺释放。研究发现，只有当这一新异性是意外(unexpected)的时候，才会

引起多巴胺释放，这一情况类似于预期错误(Diederen & Fletcher, 2021)。

基于上述研究及整合前人模型(Schultz, 2016), 可以使用一个图清晰地表示这些概念之间的关系和各自作用的阶段(图 1)。

最后值得一提的是“不完整线索” (incomplete reminders)的概念。最近，有研究者试图使用“不完整提取线索”来整合联结记忆(例如条件性恐惧记忆)与陈述性记忆再巩固的相关研究(Sinclair & Barense, 2019)。在陈述性记忆中，研究发现使用初始学习的刺激作为线索进行提取时，线索后续内容缺失或改变，会造成先前的记忆更容易被提取后的学习内容所干扰，证明预期错误在陈述性记忆再巩固过程也发挥了关键作用。研究者认为，这两类记忆要成功使用再巩固干预，都需要在提取阶段使用不完整线索，其本质上也属于预期错误(Sinclair & Barense, 2019)。

2 预期错误在条件性恐惧记忆更新中的作用

2.1 预期错误在恐惧习得与消退中的作用

预期错误被认为是恐惧习得过程的必要因素，正是从 CS 后没有强化物到实际出现强化物这一变化，介导了恐惧学习(Furlong et al., 2010)。如上文所述，中脑导水管周围灰质(periaqueductal gray, PAG)被认为是 PE 介导的恐惧习得神经环路的组成部分。最近有研究进一步考察了腹外侧导水管周围灰质(ventrolateral periaqueductal gray, vlPAG)与恐惧条件反射形成的因果关系。在大鼠中使用威胁性的 CS 匹配固定电击和不确定线索匹配一定比例电击这两种差别化程序进行习得，而后者正是通过 PE 来进行学习的。结果发现，电击引起的反应与 SPE 变化相一致，且抑制 vlPAG 会导致后续实验中恐惧反应减少，表明 PE 是在不确定的情况下保持恐惧反应的必要条件，二者具有因果关系(Walker et al., 2020)。这一结果，也符合 PE 和学习关系的既往结论(Fernandez, Boccia, et al., 2016)。

而在消退过程中，根据 RW 模型，恐惧消退中预期的CS-US和实际发现的CS-US的不匹配越大，新学习出现的可能性越大(Rescorla & Wagner, 1972)。在消退学习中，不跟随 US 的 CS+的出现，会造成两种可能的结果：恐惧激活和恐惧消退。一次或最初的CS+呈现仅对原恐惧记忆进行提取，个体表现出恐惧反应；而当多次重复出现 CS+均不匹配 US 时，就会形成抑制性记忆联结，原有的恐惧反应就会降低。这一过程中被试对预期的调整起了关键作用。预期和结果不匹配的出现促使个体有了建立新的记忆联结的需求，因此可以认为是 PE 引发了最初的恐惧消退，负性预期错误是成功消退的缘起。

近年来 Gershman 和 Monfils (2017)等人提出的基于结构学习(structure learning)机制的记忆修改模型认为，标准的恐惧习得和消退导致了两种记忆联结的形成：CS-US 和 US-no US。其中 PE 有两种作用：一方面作为一种联结性学习的信号指导个体调整 CS-US 的联结权重(weights)，另一方面作为分割信号指示一个新的潜在原因(latent cause)在何时开始活跃。在习得过程中形成的 CS-US 关系的预期使得动物在消退学习中会经历对预期的违背，即产生 PE。产生的 PE 可以通过两种途径来减弱，一是通过对原有 CS-US 联结的去学习(unlearning)或遗忘，二是将消退试次分配到一个新的潜在原因上。研究者认为，在 PE 产生之初，对少量潜在原因的简单偏向有利于遗忘；而随着消退进程的加深，PE 的持续累积最终导致一个新的潜在原因升高，最终产生了 CS-no US 联结(Gershman et al., 2017)。

然而目前无论是动物还是人类模型，对 PE 在消退学习中的神经机制的研究还较少，一般认为这一过程涉及到的主要脑区在海马(hippocampus, HIP)、腹侧被盖区(ventral tegmental area, VTA)。尤其是 HIP-VTA 回路在基于强化的记忆编码中起到了重要作用，但是这一环路在人类恐惧消退中的作用还有待证实(Sevenster et al., 2018)。另外，预期错误在恐惧消退中的作用受到诸多因素的调节，目前已知的包括睡眠尤其是快速眼动睡眠的作用，以及压力荷尔蒙的作用等。在神经递质方面，大量研究证明了中脑多巴胺神经元的活跃程度代表了实际结果比预期更好或更差的程度(Schultz, 2016; Schultz et al., 1997)，而恐惧记忆消退中负性 US (如电击)的缺失可以视为一种 “比预期更好”的结果(Raczka et al., 2011; Thiele et al., 2021)。

2.2 预期错误在恐惧记忆再巩固中的作用

2.2.1 预期错误是记忆去稳定的重要边界条件

记忆再巩固理论认为，对于已进入稳定状态

的长时记忆，在使用线索进行提取激活之后，会重返不稳定状态，变得容易受到干扰，需要经历一定的过程才能重新稳定，这一阶段被称为记忆的“再巩固”。该理论指明对记忆进行修改有两个关键窗口：巩固和再巩固。一系列研究证实了再巩固这一阶段的独立性以及干扰再巩固以消除恐惧记忆的可行性(Alberini et al., 2006; Duvarci & Nader, 2004; Lee et al., 2006; Nader et al., 2000)。在记忆再巩固的行为干预范式上，Monfils 和 Schiller 等人先后在动物和人类中证明，记忆提取后实施消退训练，可有效消除恐惧记忆并且抑制恐惧返回(Monfils et al., 2009; Schiller et al., 2010)，称之为“提取−消退”范式(retrieval-extinction, RE)。一个典型的 RE 包含连续三天实验，第一天建立 CS-US 记忆; 24 小时后呈现一个 CS 进行记忆激活，10 分钟后进行消退；第三天测试恐惧复发程度(Chen et al., 2021; Schiller et al., 2010)。近年研究发现，记忆再巩固过程还可以进一步细分为去稳定(destabilization)和再次稳定(re-stabilization)两个阶段(Elsey & Kindt, 2017; Faliagkas et al., 2018)。只有通过提取使得先前的记忆再次变得不稳定，才具有更新(消除或改写)的可能性，这一过程称为去稳定，进入这一状态的条件称为记忆再巩固的边界条件(Zuccolo & Hunziker, 2019)。因此，记忆唤醒和打开恐惧记忆再巩固窗口是保证恐惧消退效果、阻止恐惧返回的两个必不可少的前提条件。

预期错误作为一种对预期的违背(violation of expectation)，长久以来在奖赏学习(reward learning)领域得到了深入细致的研究；但在记忆再巩固理论框架内，直到近年才发现记忆提取阶段的 PE 对于开启再巩固的特殊意义。2009 年研究者 Lee 撰文指出，在记忆再巩固过程中，惊讶或者预期错误可能具有潜在的作用(Lee, 2009)。随后，Kindt 团队在 2012 年前后的一系列研究，首先在人类被试中证明，条件性恐惧记忆激活时的预期错误是记忆进入不稳定状态的必要条件(Sevenster et al., 2012, 2013, 2014)，引发了国际上对 PE 的高度关注和后续研究。目前已证明多种形式的 PE 都可以成功开启记忆再巩固，包括 TD、学习规则的 PE、US 次数的 PE 等(Chen et al., 2020; Diaz-Mataix et al., 2013; Li et al., 2019; Li et al., 2017; Sevenster et al., 2013, 2014)。PE 的作用在人类和动物被试中得到了类似的结论，并在成瘾记忆、陈述性记忆等不同记忆类型中得到验证，因而被认为是一种记忆更新过程的共通性成分(Fernandez, Bavassi, et al., 2016; Forcato et al., 2007) (Das et al., 2018; Sinclair & Barense, 2019)。新学习的启动必须以出现新异性信息为驱动力，当环境中没有有意义的新信息存在的情况下，记忆神经元的突触可塑性保持在关闭状态，不会产生新的树突及神经元之间的联系；只有当出现新信息并且与个体的适应相关、具有生存意义的时候，神经元的突触可塑性才能被激发，产生新的树突以及出现神经元之间联结强度的变化，这一活跃状态需要新的蛋白质合成才能够再次稳定下来(Diaz-Mataix et al., 2013)。

Sevenster 等人在 2013 年首先报告可以使用预期错误来引发记忆再巩固，在这一研究中，PE 被定义为恐惧习得阶段和提取阶段内容的不匹配，在提取阶段分别形成无预期错误(no PE)、正性预期错误(Positive PE)、负性预期错误(Negative PE)三种条件。其中无预期错误(no PE)是指记忆提取阶段实际出现的 CS-US 匹配关系符合预期，即与习得阶段的情况一致。进而利用心得安(propranolol)来验证是否经历了记忆再巩固，结果发现两个 PE 组都经历了记忆再巩固，而无 PE 组则没有进入再巩固(Sevenster et al., 2013)。为进一步检验预期错误在记忆再巩固中的作用，研究者们使用不同的 PE 产生方式在 RE 范式中进行考察。Díaz-Mataix 等人使用 TD 进行提取，比较第三天记忆测试的恐惧复发情况的差异(Diaz-Mataix et al., 2013)，发现 TD 组成功地通过提取消退抑制了恐惧复发，而无 TD 组则有明显的恐惧返回，从而证明了(时间性的)预期错误是启动记忆再巩固的必要条件。

在前人研究的基础上，我们在 RE 范式中使用复合刺激模型建立条件性恐惧，验证了正性、负性 PE 在复合恐惧记忆中激活记忆的效果(陈伟等, 2018)，以及对比了提取阶段 CS 的新异性(CS novelty)和 CS-US 的新异性(PE)在开启记忆再巩固上的区别(Li et al., 2019)。结果发现，正性 PE、负性 PE 在开启再巩固过程上起到的作用是一致的，实际出现的 US 大于或小于原有 US，都可以引发个体更新原始 CS-US 联结的需求；但仅有 CS 新异性不足以开启记忆再巩固，CS-US 新异性

即预期错误是激活记忆进入再巩固的关键因素和必要条件。

2.2.2 预期错误的量(degree)决定了记忆能否进入不稳定状态

PE 既是恐惧消退过程的关键因素，也是记忆进入再巩固的必要条件。然而这两个过程(记忆更新和新学习)是截然相对的，记忆提取的操作在特定的实验设置下可能引起二者之一，或二者的中间过程(limbo state) (Faliagkas et al., 2018)，因而包含 PE 的记忆提取能否引发再巩固，成为一个关键的问题。研究表明，PE 对于再巩固的触发属于必要不充分条件。在提取包含 PE 的情况下，能否成功引发记忆去稳定，还涉及到 PE 的量(degree)的问题。Sevenster 和 Kindt 等人的研究中，根据 PE 的量不同创设了三种记忆提取条件：不包含 PE 的提取(no PE)、包含单个 PE 的提取(single PE)、包含多重 PE 的提取(multiple PE)，探究其对于开启记忆再巩固的效果。结果发现，无 PE 和多重 PE 条件下，第三天恐惧测试中均有明显复发；只有单个 PE 提取条件下，没有出现恐惧复发(Sevenster et al., 2014)。该研究首次将 PE 的量作为衡量记忆是否能够进入再巩固的重要因素加以研究，对后续研究具有重要启发。我们进一步将这一范式由药物干预迁移到行为干预，得到了和药物干预相一致的结果(陈伟 等, 2018)。近年来，在陈述性记忆领域有研究使用自信心评分结合反馈的方式量化 PE，取得了一定的进展(Pine et al., 2018)。但目前在条件性恐惧的研究中，PE 的量化仍停留在较为粗略的分类层面上。也是从这些研究开始，从质性到量化考察预期错误在记忆更新中的精细作用，成为此类研究进展的重要标志。

2.2.3 记忆去稳定所需预期错误的量与原始记忆强度有关

近年来，研究者们开始注意到记忆再巩固的两类边界条件不是各自为阵的，而很可能是共同影响记忆提取的效果，因此记忆本身边界(如记忆强度)与提取边界的交互作用逐渐得到关注。两类边界条件交互作用的研究不仅有利于探明边界条件影响提取消退的作用机制，同时可以为改进实验范式、探索新的临床治疗方法提供依据。PE 作为提取边界条件，其开启记忆再巩固的作用是否会因记忆强度的不同而不同？记忆去稳定所需要的 PE 的量是否因记忆强度的增大而增多？这些都是非常值得探索的问题。最近我们在人类被试中探索了不同强度的条件性恐惧记忆在记忆激活过程中是否需要不同程度的 PE 以引发记忆去稳定。研究使用 RE 范式，在习得阶段，使用可预期的(Predictable)的电击出现时刻建立 CS-可预期 US 联合以形成较弱恐惧，使用不可预期的(Unpredictable)电击出现时刻建立 CS-不可预期 US 联合以形成较强恐惧记忆(Amadi et al., 2017)。在提取阶段，考察三种提取方式：单个 PE、多重 PE 提取和单次 CS 伴两个 US 缺失提取引发记忆去稳定的效果。结果发现，使用单个 PE 的提取消退可以防止可预期 CS-US 恐惧记忆复发，但不能阻止不可预期CS-US恐惧记忆恢复，说明单个PE不足以使较强恐惧记忆去稳定。对于多重 PE 提取和单次 CS 伴两个 US 缺失提取而言，实验发现，两种方法均抑制了恐惧记忆的自发恢复，但仅在多重 PE 组中抑制了恐惧记忆的重建，提示强恐惧记忆可能需要更多的 PE 才能破坏恐惧记忆的稳定。诱发记忆不稳定所需要的提取中 PE 的量取决于恐惧记忆的强度(Chen et al., 2020)。

而在我们另一项研究中，再次考察了 PE 在不同强度恐惧记忆下的作用，并基于谷氨酸能(Glutamatergic neurons)在激活突触可塑性方面的作用，探索了较高强度恐惧记忆下提取后急性应激对于开启再巩固的可能作用。结果发现，对于一般强度恐惧记忆，单个 PE 提取后消退可显著抑制恐惧自发恢复；而单个 PE 提取不能有效激活较强恐惧记忆，在第三天测试时会有明显的复发现象。而如果在提取后施加外源性急性应激，会进一步增大恐惧复发。研究结果体现了 PE 作为边界条件的变动性：PE 的需求量可能根据记忆强度的变化而有所不同(李俊娇 等, 2021)。但由于上述两项研究均没有直接验证高强度恐惧记忆去稳定是否需要一个更大的 PE，这一点还有待未来研究继续探索。

总之，预期错误或者提取阶段的新异性信息，在记忆由稳定状态到经由突触可塑性实现的记忆不稳定状态的过程中发挥了重要的作用。提取线索 CS 之后出现的 US 大于或小于原有 US，都可以引发个体更新原始 CS-US 联结记忆的需求。但 PE 所起的具体作用，引发的是记忆去稳定、消退还是中间状态，还需要结合记忆本身边界条件确定。基于这些研究，在前人模型的基础上(Elsey &

Kindt, 2017), 我们进一步提出了提取边界条件和记忆条件相整合的恐惧记忆再巩固作用模型, 如图 2 所示。

3　预期错误在恐惧记忆更新中的神经机制

由于预期错误的重要作用, 无论是基础研究还是临床应用上都亟待一种可操作的指标来衡量 PE 的出现(陈伟 等, 2020)。研究者指出, 人类实验相比于动物研究的优势在于可以对 US 预期进行口头汇报(Sevenster et al., 2018)。对 US 出现可能性的报告——US 主观预期, 可以作为衡量 PE 出现与否的外显指标之一。但 PE 作为一种内在过程, 其外显指征依然是有待探索的难题。这一难题的解决应建立在对多个层面机制的理解之上, 然而目前 PE 在记忆再巩固过程中作用的神经机制的研究尤为缺乏(曹杨婧文 等, 2019)。需要指出的是, PE 的产生机制与作用机制是不同概念。尽管预期错误本身的神经信号以及对应的分子机制已积累了大量研究结果, 尤其是在奖赏学习领域得到了深入研究, 但是对于开启记忆再巩固的这一作用的神经机制和分子机制的研究还很少, 有待进一步阐明。

3.1　神经环路与参与脑区

3.1.1　杏仁核

前人研究表明, 由 CS、US 的时间关系变化引起的突触可塑性的改变体现在杏仁核外侧核(lateral nucleus of amygdala, LA)中。在人类和非人类哺乳动物的条件性恐惧和成瘾研究中均显示了杏仁核在探测预期错误过程中的作用。杏仁核对意外的时间性变化或未预料到的事件作出反应。Belova 等人 2007 年在哺乳动物中的研究表明, 杏仁核的神经元会在强化学习中对意外出现的奖赏或负性刺激做出反应, 而意料之中的刺激则不会。而且可能存在不同的杏仁核神经元族群, 一类对引发唤醒和注意的意外刺激本身反应, 一类专门对于行为的效价(奖赏或惩罚)作出反应(Belova et al., 2007)。

另一项与惊讶或 PE 相关的研究, 在建立了光线−声音的联结性学习之后, 使用免疫化学方法中 Fos 蛋白(即刻早期基因 c-fos 基因的表达产物, 与学习过程密切相关)表达作为因变量, 追踪

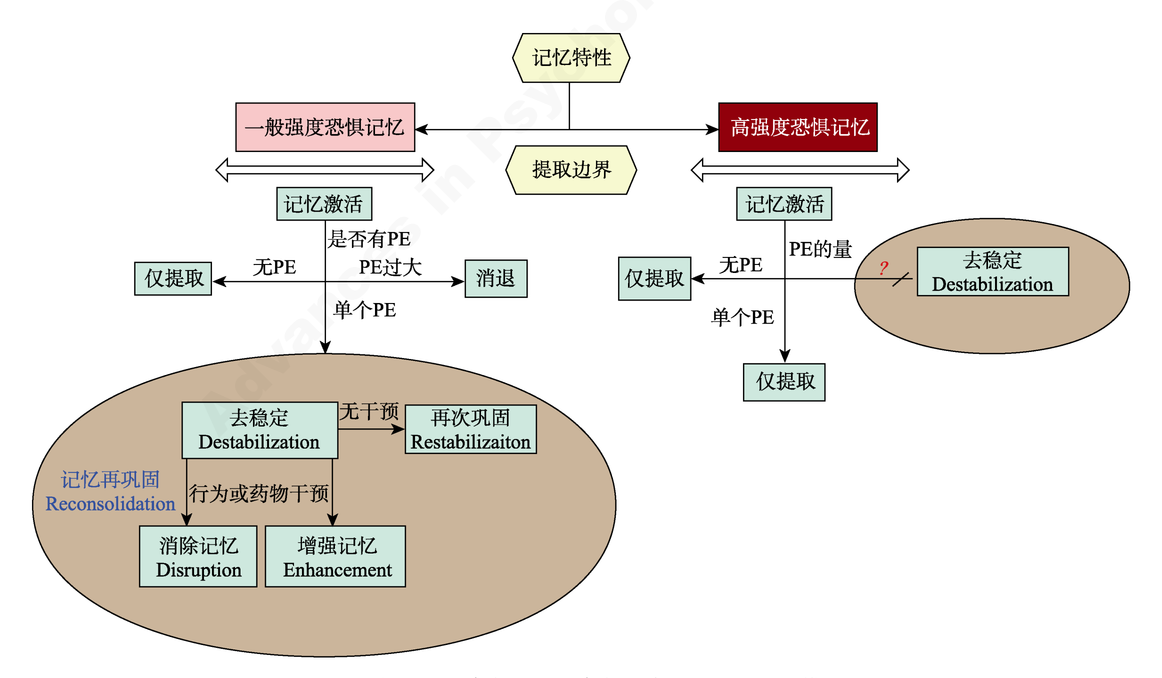


图 2　提取边界条件与记忆条件整合的记忆再巩固模型

注: 恐惧记忆本身特性(如强度)与记忆提取的边界条件(如预期错误)存在交互作用, 在一般强度恐惧记忆中, 单个 PE 可以引起记忆激活进入再巩固, 进而采用干预手段进行记忆的改写; 而 PE 过大则会引起消退过程。然而在高强度恐惧记忆中, 单个 PE 无法激活原始记忆引发去稳定, 激活记忆进入再巩固状态需要的 PE 的量目前大部分是未知的。

其在惊讶组(光线后不再固定地预测声音)和一致组(光线刺激始终可以预测声音)中的表达发现：在惊讶出现的时刻，在中央杏仁核(central nucleus of the amygdala, CN)中检测到 Fos 蛋白表达增加，在一致组检测到在基底外侧杏仁核(basolateral amygdala, BLA)中的 Fos 蛋白表达增加。这也提示杏仁核的不同区域可能在 PE 的作用机制中具有不同作用(Bucci & Macleod, 2007)。

3.1.2 腹外侧导水管周围灰质(vlPAG)

以往研究表明，导水管周围灰质(PAG)参与了预期错误计算的过程，其神经活动与 PE 的大小相一致(Roy et al., 2014)。最近又有研究使用有标记的预期错误(SPE)考察了腹外侧导水管周围灰质(ventrolateral periaqueductal gray, vlPAG)在不确定性驱动的恐惧记忆更新中的作用。对大鼠的研究表明，恐惧辨别中 vlPAG 区域单个神经元的活动与 SPE 一致并更新了后续的恐惧记忆。抑制 vlPAG 的操作可以减弱后续对不确定性线索而非威胁性线索的恐惧。该研究提示 vlPAG 可能是 SPE 计算的中心，并与恐惧记忆更新具有因果关系(Walker et al., 2020)。

关于 vlPAG 的作用方式，研究者认为很可能跟中脑 VTA 有关，研究发现 vlPAG 会发送兴奋性或抑制性的信号到 VTA，VTA 作为已知的动机与强化的加工关键区域，接受来自 vlPAG 的 γ-氨基丁酸(γ-aminobutyric acid, GABA)或谷氨酸递质输入，产生正性、负性 PE 信号，并进一步引发行为改变(Waung et al., 2019)。

3.1.3 海马

基于海马在记忆提取过程的重要作用，其在记忆再巩固中的潜在功能受到关注。在一些其他类型的记忆例如认知任务的内隐或外显联结记忆的脑成像研究中，证明了海马的作用(Duncan et al., 2009; Kumaran & Maguire, 2006; Long et al., 2016)。而对于条件性恐惧中由 CS 呈现后 US 的缺失带来的负性预期错误的神经信号，被认为同样包括海马(Spoormaker et al., 2011)。

新技术助推了神经机制的研究，在啮齿类动物中使用光遗传技术，可以精确地打开或沉默某些特定的海马神经元来研究其在恐惧条件化学习中的作用。研究者通过光遗传技术直接激活海马的齿状回(dentate gyrus, DG)和 CA1 区域，在小鼠中成功地建立起虚假的背景性恐惧记忆(Ramirez et al., 2013)，证明海马在记忆激活中的作用。同样利用光遗传手段，研究者发现海马与新皮层之间存在紧密关联，沉默海马神经元引起的记忆提取失败可以通过激活新皮层神经元来恢复(Cowansage et al., 2014)，说明在记忆提取过程中海马和新皮层具有交互作用。研究者推测，海马在处理抽象的、高阶的预期错误中具有主要作用；推测在 CA3 区产生预期结果并输送到 CA1 区和来自新皮层的感觉信息输入进行比较，进而在海马产生匹配(match)和不匹配(mismatch)的信号(Vinogradova, 2001)，而后者就是 PE。

因此研究者认为，海马在记忆提取和在线预期生成中均有参与，而海马−新皮层的交互作用在记忆提取和预期过程中是截然不同的。通过 PE 的作用，各类神经递质在决定是记忆提取还是预期更新占主导优势地位的过程中发挥了关键作用(Barron et al., 2020)。

3.1.4 前额叶皮质

前额叶皮质(prefrontal cortex, PFC)在恐惧记忆加工中始终扮演着重要角色，前额叶皮质的不同亚区参与了恐惧记忆习得、巩固、消退等各个阶段。但 PFC 在传统消退和提取消退过程中的作用存在显著差异。一项以时间性的预期错误(TD)为模式的人类 fMRI 研究发现，由预期结果缺失导致的负性 PE 信号与腹内侧前额叶(ventromedial prefrontal cortex, vmPFC)、背外侧前额叶(dorsolateral prefrontal cortex, dlPFC)、左侧眶回(left orbitofrontal cortex, LOFC)等的激活有关(Spoormaker et al., 2011)。而在 RE 范式中，最近我们对人类被试进行的恐惧记忆消退的 fMRI 研究表明，在记忆提取阶段使用 PE 进行提取，显著造成了之后的消退过程和不包含 PE 提取的消退过程的脑激活模式差异，具体表现为前者的消退过程中减少了颞下回(inferior temporal gyrus, IT)和背外侧前额叶(dlPFC)脑区的激活，并显著减少了 dlPFC-ACC (anterior cingulate cortex)、IT-dlPFC 的功能联结(Li et al., 2019)。而之前另一项未明确操纵 PE 的人类恐惧记忆提取消退 fMRI 研究，则得到提取消退和标准消退的机制区别是前者显著减少了腹内侧前额叶(vmPFC)的参与(Schiller et al., 2013)。在该研究中提取试次为单个 CS 不跟随 US，没有具体操控 PE，但由于伴随的 US 缺失，或可视为一种隐含的 PE。可见，前额叶皮层及其亚区在 PE 开

启再巩固中扮演了重要角色, 但目前人类研究仍较少, 其发挥的具体作用还有待研究的进一步发掘。

基于我们在该领域中行为和脑成像层面的人类研究和参考前人研究结果, 我们提出了前额叶皮质及相关脑区在人类预期错误驱动记忆更新过程的作用模型(图 3)。

3.2　环路中的神经调节

各种神经递质(neurotransmitter)与 PE 的作用有着密切关联, 基于错误的学习受到神经系统中特定神经递质的重要调节, 其中最显著的就是多巴胺能。多巴胺神经元由于会对意外结果作出直接的显著反应, 而被认为与记忆更新和去稳定有关。中脑多巴胺会在正性 PE 条件下增多, 在无 PE 条件下没有变化, 以及在负性 PE 情况下减少。和中脑 VTA 中的多巴胺消耗的模式不同, 在 BLA 中对 PE 的反应是不分效价的 (Schultz, 2016)。

近年来研究还揭示, 中脑多巴胺除了会对意外出现的奖赏刺激作出反应之外, 还会对惩罚刺激的意外消失进行反应, 从而在恐惧记忆消除的研究以及以此为基础的暴露疗法改进上提供了重要启示(Hernandez et al., 2018)。Papalini 和 Beckers 等人(2020)提出, PE 及其引发的中脑多巴胺对恐惧记忆消退的影响, 从分子机制到行为层面的临床治疗可以分为三步: (1)作为结果的负性刺激(US)的缺失引起 PE 在中脑纹状体的伏隔核(nucleus accumbens, NAcc) /VTA 区域引发多巴胺燃烧; (2)该DA信号驱动新的安全记忆产生, 这一过程可能主要涉及 D2R (G 蛋白偶联受体(G protein-coupled receptor, GPCR)中的 D2 类受体)介导的多巴胺能信号。DA 从中脑传输到前额叶皮质, 一方面在 vmPFC 更新对威胁的负性预期, 另一方面可能在外侧前额叶(lateral prefrontal cortex, lPFC)实现消退记忆的提取, 这一过程可能涉及了 PFC 和海马中的 D1R (GPCR 的 D1 类受体)介导的多巴胺能信号通路。(3)在行为层面上, 需进一步考察基于多巴胺能干预的多种操作在促进消退和保持暴露疗法效果上的作用(Papalini et al., 2020)。

除此之外, 谷氨酸(glutamate, Glu)神经递质在 PE 开启再巩固过程的作用值得关注。早在 2006 年即有研究者提出在成瘾记忆中, PE 引起的 DA 和 Glu 可能是同步释放的(corelease), 称为 Glu-DA 共同传输(Glu-DA co-transmission) (Lapish et al., 2006), 因此这两种神经递质可能是共同调节 PE 的作用, 但是目前这方面研究比较缺乏。另一方面, 谷氨酸能在记忆去稳定过程中的作用近年来已逐步得到揭示。分子层面的研究表明, 记忆能否去稳定, 取决于一种关键性的神经递质——Glu 的离子型受体NMDA(N-甲基-D-天冬氨酸, N-methyl-D-aspartate)亚单位 GluN2B 的含量; 尤其是 GluN2B/GluN2A 的比例, 决定了恐惧记忆能否进入不稳定状态(Milton et al., 2013; Shipton & Paulsen, 2014)。最近一项研究发现, 在强恐惧记忆条件下增加 BLA 中 GluN2B 的水平, 可以修改原本由于强度导致的抵抗再巩固干预的记忆(de Solis et al., 2019)。由于目前 PE 引发记忆去稳定

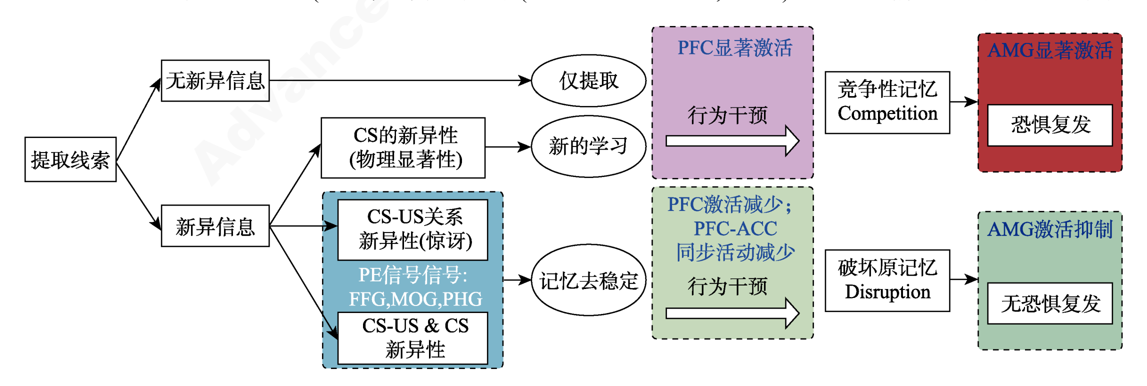


图 3　前额叶皮质及相关脑区在预期错误驱动记忆更新过程中的作用

注:在提取边界条件上, 提取阶段是否存在 CS-US 关系的新异性是进入再巩固的必要条件, 而仅有 CS 的新异性则引发新的消退学习。消退学习即传统消退与提取消退在激活脑区上存在显著差别, 集中体现在前额叶皮质尤其是背外侧前额叶(dlPFC)区域。提取消退过程中 dlPFC 和颞下回的激活显著降低, 同时 dlPFC 与前扣带回等脑区的功能联结显著下降。相反在传统消退过程中 dlPFC 有显著激活。

的分子机制仍然未知, 因而从此类研究中我们可以推测其作用的神经环路也与谷氨酸能密切相关。

4 预期错误在恐惧记忆更新研究中的未来趋势与方向

4.1 基于预期错误计算模型的量化研究

如前文所述, 近年来预期错误在记忆再巩固中的作用研究的主要进展之一就是从质性到量化, 从 PE 的有或无到 PE 不同量的作用, 研究逐步精细化; 因此 PE 量化成为目前该领域一个主要问题。我们认为, 基于预期错误计算模型的量化研究可以借鉴以下方面的研究进展:

(1)其他记忆类型中量化 PE 和记忆更新关系的研究。Pine 等人首先在陈述性记忆中使用“基于回忆的选择−信心评分−反馈”范式, 在行为和脑成像层面探索了纹状体区域的 PE 的量驱动记忆更新的机制(Pine et al., 2018)。我们认为, 恐惧记忆中 PE 操纵和量化可以借鉴陈述性记忆中的成功范式, 在不同层面上考察 PE 的量与恐惧记忆更新的关系。

(2)动物模型中精神病相关机制的研究。最近有研究将精神分裂症的主要症状之一——幻觉进行操作化, 借由心理物理学中的信号检测论将幻觉操作性定义为: 高度自信的虚报(high confidence false alarm)。该研究运用包含预期错误的信念更新(belief update)公式, 模拟计算了 PE、学习能力和信念更新的量, 并在动物和人类中加以验证(Schmack et al., 2021)。此类精神病模型中的 PE 计算方法, 也对人类恐惧记忆中 PE 的量化具有高度启示。

(3)新的预期与学习理论模型的研究。近年来在学习领域持续涌现出新的模型, 比传统模型能更灵活地解释新的学习现象。Osan 等人在 2011 年提出基于信息不匹配的记忆再巩固和消退的神经网络模型认为, 赫布学习(Hebbian learning)产生突触强化, 预期错误导致突触退化, 两者的平衡与否形成了突触权重。当感觉输入信号与记忆原始痕迹略有不同时, 原始记忆会更新突触权重的配置。而如果输入信号与记忆原始印迹显著不同, 则会产生新的记忆痕迹与前者对抗。记忆再巩固过程, 则是由神经网络输入信号(与原始模式相似但不完全相同)和回忆引发的恐惧记忆之间的差异引起的, 以预测错误为特征(Osan et al., 2011), 进而提出了 PE 的计算模型。这一模型得到了后人研究的数据支持, 可用以预测行为结果, 从而使得对边界条件的实验性测量成为可能(Radiske et al., 2017)。

因此未来研究应借鉴上述三个方面的研究进展, 在恐惧记忆再巩固框架内使用 PE 的计算模型, 建立起 PE 和记忆消除更精确的定量关系, 以加深对其作用的基本理解。

4.2 探索预期错误与其他边界条件的交互作用及相关指标

预期错误作为提取边界条件, 和记忆本身特性存在相互影响(图 2), 但当前研究中少见直接将两者相结合的研究, 尤其是人类研究。提取边界与记忆本身特性(如强度)结合的研究是未来该领域一个重要方向, 这对于加深对 PE 作用机制的认识以及促进该范式的临床应用都具有重要意义。

这也意味着, 应致力于探索可行的记忆再巩固指标问题。记忆再巩固研究挥之不去的一个问题就是缺乏相应的外显指标, 目前尚没有研究在真正意义上找到一个可以指征记忆进入再巩固的指标。以预期错误为例, 由于其是记忆去稳定的必要非充分条件, 因此不具备成为指标的条件; 在临床实际中也很难判断患者是否出现了 PE 及其大小。再巩固指标的缺乏也使得目前仍有部分研究者质疑这一阶段的存在(Miller & Matzel, 2006), 他们认为记忆再巩固不是对于实验结果的唯一解释, 甚至直接使用传统的记忆编码加工理论就可以解释实验结果。基于对于各种记忆类型再巩固的大量神经生物学证据, 我们认为记忆再巩固阶段是客观存在的, 但仍需要继续探索可以指征的外显指标, 包括 PE 的相应指标。

4.3 考察不同类型显著性在记忆再巩固中的作用

前文总结了不同类型显著性之间的关系, 主要从记忆加工的不同深度上加以区分(图 1)。PE 作为一种显著性信息, 其重点在于行为的结果与行为的关系上, 而另一类新异性信息 CS 的显著性(stimulus salience)也得到了逐渐增多的关注。刺激物本身的新异性和刺激−结果联结的 PE 在提取干预范式中的作用有什么区别, 二者分别受到什么因素的影响及在神经机制层面上对应了何种通路等, 应当加以研究。

前人研究表明, 虽然刺激物的物理变化会增强 DA, 却不足以引起 DA 的释放, 刺激物本身的

变化要配合强化物才具有动力性(Schultz, 2016)。但是由于刺激物的变化会调动注意系统, 产生更多的去甲肾上腺素(Norepinephrine, NE)以及引起朝向反射, 因此也会对提取干预过程产生影响(Li et al., 2017)。当刺激物变化叠加 PE 时比单独呈现 PE 有怎样的效果变化, 可以进一步加以验证。而在神经机制层面, 有研究发现中脑纹状体的不同亚区的 DA 活动反映不同类型显著性: 腹侧纹状体(ventral striatum, VS)主要加工奖赏预期(reward expectation), 而纹状体尾部(tail of striatum, TS)主要加工知觉预期(perceptual expectation) (Schmack et al., 2021)。这些研究证明, 不同类型显著性的机制很可能是不同的, 影响了行为的不同方面。因此未来研究的一个关注点是各类型显著性对人类恐惧记忆提取干预的影响及机制。

4.4 预期错误作用的个体差异研究

预期错误驱动恐惧更新的作用模式可能存在个体差异, 然而目前这类研究尤其缺乏。在个体差异中, 临床和非临床被试的差异是很重要的一类, 其差异主要包括以下方面:

首先, 异常人群可能存在 PE 信号功能失调。研究表明, 临床抑郁被试虽然和非抑郁被试在强化学习上没有显著差异, 但其在场景分类的效价评估任务中表现出更大的错误。两者在记忆测试中没有差异, 但抑郁被试有更大的负性 PE, 负性 PE 比正性 PE 更能增强其情景记忆; 而非抑郁被试有更大的正性 PE (Rouhani & Niv, 2019)。Yaple 和 Yu 等人的研究也发现, 在精神分裂症和抑郁症人群中存在 PE 加工异常, 这两类临床被试在 PE 的脑区信号上存在后扣带回区域活动的一致性, 而在健康被试中没有这一效应(Yaple et al., 2021)。

其次, 奖赏学习中奖赏加工失调。快感缺失是抑郁症的核心特征, 奖赏学习中奖赏加工失调是抑郁症的发病机制之一, 其中包括预期调节功能失调。抑郁症患者的奖赏性 PE 量显著小于正常人, 正性效价对于抑郁症人群的强化作用明显更小。在因果关系中, 研究者认为, 预期错误及其调节系统可能参与了精神疾病的形成过程(Beckers & Kindt, 2017)。因而, 进一步提高对抑郁症中奖赏加工功能障碍的理解有望改进抑郁症的诊断和治疗(Admon & Pizzagalli, 2015), 在未来需要对包括抑郁症在内的临床被试进行 PE 功能的研究。

此外, 临床被试与正常人群相比还存在精神疾病本身伴随或引发的压力状况差异。以 PTSD 为例, 在记忆唤起或提取过程也伴随着压力的升高, 被试处于高压状态。而压力会造成记忆去稳定的边界条件, 制约 PE 激活原始记忆的效果。在抑郁症的动物模型和人类研究中均证实, 压力是初始抑郁发作的常见触发因素, 慢性压力可以抑制海马神经活动发生, 抑制中脑边缘多巴胺神经元, 并使杏仁核对负面信息的反应敏感。这些机制也可以解释抑郁成人中对积极材料的记忆中断和对消极材料的记忆增强现象(Dillon & Pizzagalli, 2018)。

因此, 未来应注重临床被试和正常人群的比较研究, 以及在研究中运用精神障碍相关的记忆模型, 考虑压力等一些急性、慢性因素的影响, 才能使得基础研究的结论有更好的临床指向性。

4.5 使用多学科手段探索 PE 在记忆更新中作用的神经与分子机制

更重要的问题仍然是, 预期错误到底是如何开启记忆再巩固的? 其背后的神经生理机制是怎样的? 这可以分解为两方面问题: PE 本身的神经信号, PE 作用的过程机制。对后者而言, 单个 CS 提取在引发再巩固和未引发再巩固之间在神经过程上有什么区别, PE 的作用主要体现在哪个时间区间上等问题, 目前还没有发现相应的研究。由于找到记忆再巩固的外显指标对于该范式的应用至关重要, 因此与其密切相关的 PE 作用机制的澄清就显得非常关键。未来应致力于从多个层面上继续探索预期错误打开记忆再巩固过程的机制。

而在分子机制上, 多类经典神经递质在记忆去稳定过程也发挥了重要的作用, 除 DA 以外, 还包括 NE, 乙酰胆碱(Acetylcholine, Ach)、五羟色胺(5-hydroxytryptamine, 5-HT)、GABA, 以及脑源性神经营养因子(Brain-derived neurotrophic factor, BNDF)等, 相关进展可参见 Wideman 等人关于神经递质在记忆去稳定过程中的作用综述(Wideman et al., 2018)。PE 和这些神经递质存在怎样的关系, 结合 PE 与其他边界条件之间的交互作用如何对记忆更新发挥影响, 也是值得进一步探索的问题。

记忆研究在各个层面上不是割裂的, 而是相互补充相互解释的。由于生物技术的直接性, 可以对行为层面上无法直接触及的因素进行研究。例如光遗传学将光敏离子通道(光感蛋白)特异性表达在特定类型神经细胞上, 可以通过光刺激改

变这些神经细胞的放电模式，研究其功能。目前PE的内在机制仍有许多部分是未知的，绝大多数关于 PE 的研究仍建立在通过外部和内部表征之间的不匹配来驱动记忆更新这一模型之上(Fernandez, Boccia, et al., 2016)。有研究者断言，未来非侵入性脑刺激(Non-invasive Brain Stimulation, NIBS)，包括经颅磁刺激(Transcranial Magnetic Stimulation, TMS)与经颅电刺激(Transcranial Electric Stimulation, tES)以及光遗传学等，将从恐惧记忆的神经生物学方面进行更具体的操作，从而为由创伤、压力和焦虑引起的病理性恐惧指明具体的治疗目标(Borgomaneri et al., 2021)。我们相信，多学科联合和多种技术手段运用是未来在包括 PE 更新恐惧记忆在内的一系列基础研究及临床转化研究的必然趋势。

总的来说，作为学习和决策领域的经典主题，预期错误在近年来兴起的记忆再巩固理论中又重新焕发了生机。其意义主要是由于以恐惧记忆为特征的多种临床精神障碍的治疗，在记忆再巩固的研究中看到了彻底更新或进一步消除的希望。作为开启记忆再巩固状态的至关重要的变量之一，预期错误在未来必将继续得到深入研究并在临床应用上发挥出更大的作用。

参考文献

曹杨婧文，李俊娇，陈伟，杨勇，胡琰健，郑希付. (2019). 条件性恐惧记忆消退的提取干预范式及其作用的神经机制. *心理科学进展, 27*(2), 268−277. doi: 10.3724/sp.J.1042.2019.00268

陈伟，李俊娇，曹杨婧文，杨勇，胡琰健，郑希付. (2018). 预期错误在复合恐惧记忆提取消退中的作用. *心理学报, 50*(7), 739–749. doi: 10.3724/sp.J.1041.2018.00739

陈伟，李俊娇，林小裔，张晓霞，郑希付. (2020). 行为干预情绪记忆再巩固:从实验室到临床转化. *心理科学进展, 28*(2), 240−251. doi: 10.3724/sp.J.1042.2020.00240

李俊娇，陈伟，胡琰健，曹杨婧文，郑希付. (2021). 预期错误与急性应激对不同强度恐惧记忆提取消退的影响. *心理学报, 53*(6), 587−602. https://doi.org/10.3724/SP.J.1041.2021.00587

Admon, R., & Pizzagalli, D. A. (2015). Dysfunctional reward processing in depression. *Current Opinion in Psychology, 4*, 114–118. doi: 10.1016/j.copsyc.2014.12.011

Alberini, C. M., Milekic, M. H., & Tronel, S. (2006). Mechanisms of memory stabilization and de-stabilization. *Cellular & Molecular Life Sciences, 63*(9), 999–1008. doi: 10.1007/s00018-006-6025–7

Amadi, U., Lim, S. H., Liu, E., Baratta, M. V., & Goosens, K. A. (2017). Hippocampal processing of ambiguity enhances fear memory. *Psychological Science, 28*(2), 143–161. doi: 10.1177/0956797616674055

Barron, H. C., Auksztulewicz, R., & Friston, K. (2020). Prediction and memory: A predictive coding account. *Progress in Neurobiology, 192*, 101821. doi: 10.1016/j.pneurobio.2020.101821

Beckers, T., & Kindt, M. (2017). Memory reconsolidation interference as an emerging treatment for emotional disorders: Strengths, limitations, challenges, and opportunities. *Annual Review of Clinical Psychology, 13*, 99–121. doi: 10.1146/annurev-clinpsy-032816-045209

Belova, M. A., Paton, J. J., Morrison, S. E., & Salzman, C. D. (2007). Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron, 55*(6), 970–984. doi: 10.1016/j.neuron.2007.08.004

Borgomaneri, S., Battaglia, S., Sciamanna, G., Tortora, F., & Laricchiuta, D. (2021). Memories are not written in stone: Re-writing fear memories by means of non-invasive brain stimulation and optogenetic manipulations. *Neuroscience and Biobehavioral Reviews, 127*, 334–352. doi: 10.1016/j.neubiorev.2021.04.036

Bucci, D. J., & Macleod, J. E. (2007). Changes in neural activity associated with a surprising change in the predictive validity of a conditioned stimulus. *European Journal of Neuroscience, 26*(9), 2669–2676. doi: 10.1111/j.1460-9568.2007.05902.x

Chen, W., Li, J., Xu, L., Zhao, S., Fan, M., & Zheng, X. (2020). Destabilizing different strengths of fear memories requires different degrees of prediction error during retrieval. *Frontiers in Behavioral Neuroscience, 14*, 598924. doi: 10.3389/fnbeh.2020.598924

Chen, W., Li, J., Zhang, X., Dong, Y., Shi, P., Luo, P., & Zheng, X. (2021). Retrieval-extinction as a reconsolidation-based treatment for emotional disorders:Evidence from an extinction retention test shortly after intervention. *Behaviour Research and Therapy, 139*, 103831. doi: 10.1016/j.brat.2021.103831

Colombo, M. (2014). Deep and beautiful. The reward prediction error hypothesis of dopamine. *Studies in History and Philosophy of Biological and Biomedical Sciences, 45*, 57–67. doi: 10.1016/j.shpsc.2013.10.006

Cowansage, K. K., Shuman, T., Dillingham, B. C., Chang, A., Golshani, P., & Mayford, M. (2014). Direct reactivation of a coherent neocortical memory of context. *Neuron, 84*(2), 432–441. doi: 10.1016/j.neuron.2014.09.022

Das, R. K., Gale, G., Hennessy, V., & Kamboj, S. K. (2018). A prediction error-driven retrieval procedure for destabilizing and rewriting maladaptive reward memories in hazardous

drinkers. *Journal of Visualized Experiments,* (131), e56097. doi: 10.3791/56097

de Solis, C. A., Gonzalez, C. U., Galdamez, M. A., Perish, J. M., Woodard, S. W., Salinas, C. E., … Ploski, J. E. (2019). Increasing synaptic glun2b levels within the basal and lateral amygdala enables the modification of strong reconsolidation resistant fear memories. *bioRxiv*, 537142. doi:10.1101/537142

Diaz-Mataix, L., Ruiz Martinez, R. C., Schafe, G. E., LeDoux, J. E., & Doyere, V. (2013). Detection of a temporal error triggers reconsolidation of amygdala-dependent memories. *Curre* 72. doi: 10.1016/j.cub.2013.01.053

Diederen, K. M. J., & Fletcher, P. C. (2021). Dopamine, prediction error and beyond. *The Neuroscientist, 27*(1), 30–46. doi: 10.1177/1073858420907591

Dillon, D. G., & Pizzagalli, D. A. (2018). Mechanisms of memory disruption in depression. *Trends in Neurosciences, 41*(3), 137–149. doi: 10.1016/j.tins.2017.12.006

Duncan, K., Curtis, C., & Davachi, L. (2009). Distinct memory signatures in the hippocampus: Intentional states distinguish match and mismatch enhancement signals. *The Journal of Neuroscience, 29*(1), 131–139. doi: 10.1523/JNEUROSCI.2998-08.2009

Duvarci, S., & Nader, K. (2004). Characterization of fear memory reconsolidation. *The Journal of Neuroscience, 24*(42), 9269–9275. doi: 10.1523/JNEUROSCI.2971-04.2004

Elsey, J. W. B., & Kindt, M. (2017). Tackling maladaptive memories through reconsolidation: From neural to clinical science. *Neurobiology of Learning and Memory, 142*(Pt A), 108–117. doi: 10.1016/j.nlm.2017.03.007

Ergo, K., de Loof, E., & Verguts, T. (2020). Reward prediction error and declarative memory. *Trends in Cognitive Sciences, 24*(5), 388–397. doi: 10.1016/j.tics.2020.02.009

Faliagkas, L., Rao-Ruiz, P., & Kindt, M. (2018). Emotional memory expression is misleading: Delineating transitions between memory processes. *Current Opinion in Behavioral Sciences, 19*, 116–122. doi: 10.1016/j.cobeha.2017.12.018

Fernandez, R. S., Bavassi, L., Forcato, C., & Pedreira, M. E. (2016). The dynamic nature of the reconsolidation process and its boundary conditions: Evidence based on human tests. *Neurobiology of Learning and Memory, 130*, 202–212. doi: 10.1016/j.nlm.2016.03.001

Fernandez, R. S., Boccia, M. M., & Pedreira, M. E. (2016). The fate of memory: Reconsolidation and the case of prediction error. *Neuroscience and Biobehavioral Reviews, 68*, 423–441. doi: 10.1016/j.neubiorev.2016.06.004

Forcato, C., Burgos, V. L., Argibay, P. F., Molina, V. A., Pedreira, M. E., & Maldonado, H. (2007). Reconsolidation of declarative memory in humans. *Learning & Memory, 14*(4), 295–303. doi: 10.1101/lm.486107

Furlong, T. M., Cole, S., Hamlin, A. S., & McNally, G. P. (2010). The role of prefrontal cortex in predictive fear learning. *Behavioral Neuroscience, 124*(5), 574–586. doi: 10.1037/a0020739

Gershman, S. J., Monfils, M. H., Norman, K. A., & Niv, Y. (2017). The computational nature of memory modification. *Elife, 6*. doi: 10.7554/eLife.23763

Hernandez, X. I., Vogel, P., Betz, S., Kalisch, R., Sigurdsson, T., & Duvarci, S. (2018). Dopamine neurons drive fear extinction learning by signaling the omission of expected aversive outcomes. *Elife, 7*. doi: 10.7554/eLife.38818

Kim, H. F., Ghazizadeh, A., & Hikosaka, O. (2014). Separate groups of dopamine neurons innervate caudate head and tail encoding flexible and stable value memories. *Frontiers in Neuroanatomy, 8*, 120. doi: 10.3389/fnana.2014.00120

Kumaran, D., & Maguire, E. A. (2006). The dynamics of hippocampal activation during encoding of overlapping sequences. *Neuron, 49*(4), 617–629. doi: 10.1016/j.neuron.2005.12.024

Lapish, C. C., Seamans, J. K., & Chandler, L. J. (2006). Glutamate-dopamine cotransmission and reward processing in addiction. *Alcoholism: Clinical and Experimental Research, 30*(9), 1451–1465. doi: 10.1111/j.1530-0277.2006.00176.x

Lee, J. L. (2009). Reconsolidation: Maintaining memory relevance. *Trends in Neurosciences, 32*(8), 413–420. doi: 10.1016/j.tins.2009.05.002

Lee, J. L., Milton, A. L., & Everitt, B. J. (2006). Reconsolidation and extinction of conditioned fear: Inhibition and potentiation. *The Journal of Neuroscience, 26*(39), 10051–10056. doi: 10.1523/JNEUROSCI.2466-06.2006

Li, J., Chen, W., Caoyang, J., Hu, Y., Yang, Y., Xu, L., … Zheng, X. (2019). Role of prediction error in destabilizing fear memories in retrieval extinction and its neural mechanisms. *Cortex, 121*, 292–307. doi: 10.1016/j.cortex.2019.09.003

Li, J., Chen, W., Caoyang, J., Wu, W., Jie, J., Xu, L., & Zheng, X. (2017). Moderate partially reduplicated conditioned stimuli as retrieval cue can increase effect on preventing relapse of fear to compound stimuli. *Frontiers in Human Neuroscience, 11*, 575. doi: 10.3389/fnhum.2017.00575

Long, N. M., Lee, H., & Kuhl, B. A. (2016). Hippocampal mismatch signals are modulated by the strength of neural predictions and their similarity to outcomes. *The Journal of Neuroscience, 36*(50), 12677–12687. doi: 10.1523/JNEUROSCI.1850-16.2016

McNally, G. P., Johansen, J. P., & Blair, H. T. (2011). Placing prediction into the fear circuit. *Trends in Neurosciences, 34*(6), 283–292. doi: 10.1016/j.tins.2011.03.005

Miller, R. R., & Matzel, L. D. (2006). Retrieval failure versus memory loss in experimental amnesia: Definitions and processes. *Learning & Memory, 13*(5), 491–497. doi: 10.1101/lm.241006

Milton, A. L., Merlo, E., Ratano, P., Gregory, B. L., Dumbreck, J. K., & Everitt, B. J. (2013). Double dissociation of the requirement for glun2b- and glun2a-containing nmda receptors in the destabilization and restabilization of a reconsolidating memory. *The Journal of Neuroscience, 33*(3), 1109–1115. doi: 10.1523/JNEUROSCI.3273-12.2013

Monfils, M.-H., Cowansage, K. K., Klann, E., & LeDoux, J. E. (2009). Extinction-reconsolidation boundaries: Key to persistent attenuation of fear memories. *Science, 324*(951), 951–955. doi: 10.1126/science.1167975

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fearmemories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature, 406*, 722–726.

Osan, R., Tort, A. B., & Amaral, O. B. (2011). A mismatch-based model for memory reconsolidation and extinction in attractor networks. *Plos One, 6*(8), e23113. doi: 10.1371/journal.pone.0023113

Papalini, S., Beckers, T., & Vervliet, B. (2020). Dopamine: From prediction error to psychotherapy. *Translational Psychiatry, 10*(1), 164. doi: 10.1038/s41398-020-0814-x

Pearce, J. M., & Hall, G. (1980). A model for pavlovian learning:Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87*(6), 532–552.

Pine, A., Sadeh, N., Ben-Yakov, A., Dudai, Y., & Mendelsohn, A. (2018). Knowledge acquisition is governed by striatal prediction errors. *Nature Communications, 9*(1), 1673. doi: 10.1038/s41467-018-03992-5

Raczka, K. A., Mechias, M. L., Gartmann, N., Reif, A., Deckert, J., Pessiglione, M., & Kalisch, R. (2011). Empirical support for an involvement of the mesostriatal dopamine system in human fear extinction. *Translational Psychiatry, 1*, e12. doi: 10.1038/tp.2011.10

Radiske, A., Gonzalez, M. C., Conde-Ocazionez, S. A., Feitosa, A., Kohler, C. A., Bevilaqua, L. R., & Cammarota, M. (2017). Prior learning of relevant nonaversive information is a boundary condition for avoidance memory reconsolidation in the rat hippocampus. *The Journal of Neuroscience, 37*(40), 9675–9685. doi: 10.1523/JNEUROSCI.1372-17.2017

Ramirez, S., Liu, X., Lin, P. A., Suh, J., Pignatelli, M., Redondo, R. L., … Tonegawa, S. (2013). Creating a false memory in the hippocampus. *Science, 341*(6144), 387. Retrieved from http://science.sciencemag.org/content/341/6144/387.abstract

Rescorla, R. A., and Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black and W. F. Prokasy (Eds.), *Classical Conditioning. II. Current Research and Theory* (pp.64–69). NY: Appleton-Century-Crofts.

Rouhani, N., & Niv, Y. (2019). Depressive symptoms bias the prediction-error enhancement of memory towards negative events in reinforcement learning. *Psychopharmacology, 236*(8), 2425–2435. doi: 10.1007/s00213-019-05322-z

Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G. E., & Wager, T. D. (2014). Representation of aversive prediction errors in the human periaqueductal gray. *Nature Neuroscience, 17*(11), 1607–1612. doi: 10.1038/nn.3832

Schiller, D., Kanen, J. W., LeDoux, J. E., Monfils, M. H., & Phelps, E. A. (2013). Extinction during reconsolidation of threat memory diminishes prefrontal cortex involvement. *Proceedings of the National Academy of Sciences of the United States of America, 110*(50), 20040–20045. doi: 10.1073/pnas.1320322110

Schiller, D., Monfils, M. H., Raio, C. M., Johnson, D. C., Ledoux, J. E., & Phelps, E. A. (2010). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature, 463*(7277), 49–53. doi: 10.1038/nature08637

Schmack, K., Bosc, M., Ott, T., Sturgill, J. F., & Kepecs, A. (2021). Striatal dopamine mediates hallucination-like perception in mice. *Science, 372*(6537), eabf4740. doi: 10.1126/science.abf4740

Schultz, W. (2000). Multiple reward signals in the brain. *Nature Reviews Neuroscience, 1*, 199–207. https://doi.org/10.1038/35044563

Schultz, W. (2016). Dopamine reward prediction-error signalling: A two-component response. *Nature Reviews Neuroscience, 17*(3), 183–195. doi: 10.1038/nrn.2015.26

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599.

Sevenster, D., Beckers, T., & Kindt, M. (2012). Retrieval per se is not sufficient to trigger reconsolidation of human fear memory. *Neurobiology of Learning and Memory, 97*(3), 338–345. doi: 10.1016/j.nlm.2012.01.009

Sevenster, D., Beckers, T., & Kindt, M. (2013). Prediction error governs pharmacologically induced amnesia for learned fear. *Science, 339*, 830–833. doi: 10.1126/science.1231357

Sevenster, D., Beckers, T., & Kindt, M. (2014). Prediction error demarcates the transition from retrieval, to reconsolidation, to new learning. *Learning & Memory, 21*(11), 580–584. doi: 10.1101/lm.035493.114

Sevenster, D., Visser, R. M., & D'Hooge, R. (2018). A translational perspective on neural circuits of fear extinction: Current promises and challenges. *Neurobiology of Learning and Memory, 155*, 113–126. doi: 10.1016/j.nlm.2018.07.002

Shipton, O. A., & Paulsen, O. (2014). Glun2a and glun2b subunit-containing nmda receptors in hippocampal plasticity. *Philosophical Transactions of the Royal Society B., 369*(1633), 20130163. doi: 10.1098/rstb.2013.0163

Sinclair, & Barense. (2019). Prediction error and memory reactivation: How incomplete reminders drive reconsolidation.

Trends in Neurosciences, 42(10), 727–739. doi: 10.1016/j.tins.2019.08.007

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: Prediction error influences episodic memory reconsolidation. *Learning and Memory, 25*. doi: 10.1101/lm.046912.117

Spoormaker, V. I., Andrade, K. C., Schroter, M. S., Sturm, A., Goya-Maldonado, R., Samann, P. G., & Czisch, M. (2011). The neural correlates of negative prediction error signaling in human fear conditioning. *Neuroimage, 54*(3), 2250–2256. doi: 10.1016/j.neuroimage.2010.09.042

Starkweather, C. K., Babayan, B. M., Uchida, N., & Gershman, S. J. (2017). Dopamine reward prediction errors reflect hidden-state inference across time. *Nature Neuroscience, 20*(4), 581–589. doi: 10.1038/nn.4520

Thiele, M., Yuen, K. S. L., Gerlicher, A. V. M., & Kalisch, R. (2021). A ventral striatal prediction error signal in human fear extinction learning. *NeuroImage, 229*, 117709. doi: 10.1016/j.neuroimage.2020.117709

Vinogradova, O. S. (2001). Hippocampus as comparator: Role of the two input and two output systems of the hippocampus in selection and registration of information. *Hippocampus, 11*(5), 578–598. doi: 10.1002/hipo.1073

Walker, R. A., Wright, K. M., Jhou, T. C., & McDannald, M. A. (2020). The ventrolateral periaqueductal grey updates fear via positive prediction error. *European Journal of Neuroscience, 51*(3), 866–880. doi: 10.1111/ejn.14536

Waung, M. W., Margolis, E. B., Charbit, A. R., & Fields, H. L. (2019). A midbrain circuit that mediates headache aversiveness in rats. *Cell Reports, 28*(11), 2739–2747 e2734. doi: 10.1016/j.celrep.2019.08.009

Wideman, C. E., Jardine, K. H., & Winters, B. D. (2018). Involvement of classical neurotransmitter systems in memory reconsolidation: Focus on destabilization. *Neurobiology of Learning and Memory, 156*, 68–79. doi: 10.1016/j.nlm.2018.11.001

Yaple, Z. A., Tolomeo, S., & Yu, R. (2021). Abnormal prediction error processing in schizophrenia and depression. *Human Brain Mapping*. doi: 10.1002/hbm.25453

Zuccolo, P. F., & Hunziker, M. H. L. (2019). A review of boundary conditions and variables involved in the prevention of return of fear after post-retrieval extinction. *Behavioural Processes, 162*, 39–54. doi: 10.1016/j.beproc.2019.01.011

The function and mechanisms of prediction error in updating fear memories

LI Junjiao[1], CHEN Wei[2,3,4], SHI Pei[2,3,4], DONG Yuanyuan[2,3,4], ZHENG Xifu[2,3,4]

([1] *College of Teacher Education, Guangdong University of Education, Guangzhou* 510303, *China*)

([2] *School of Psychology, South China Normal University, Guangzhou* 510631, *China*)

([3] *Center for Studies of Psychological Application, South China Normal University, Guangzhou* 510631, *China*)

([4] *Guangdong Key Laboratory of Mental Health and Cognitive Science, Guangzhou* 510631, *China*)

Abstract: According to the error-driven learning theory, the mismatch between expected outcome of behavior and actual result, known as "Prediction error" or PE, is the driving factor of new learning. Prediction error differs from other types of salience, such as physical salience, surprise, or novelty, in terms of distinct periods of information processing, as well as in its relationship with memory updating. The reconsolidation interference paradigm has been shown to be effective in neutralizing conditioned fear memory in humans, where the prediction error involved in memory reactivation is required to reactivate memory for reconsolidation. In the behavioral mechanisms of PE in promoting fear memory updating, it is found that PE is a necessary but not sufficient condition of memory destabilization. Memory reactivation must include appropriate degree of PE; however, properties of the memory must be taken into account when determining the fate of memory following reactivation, which could be destabilization, extinction or limbo. In the neural mechanism of PE in fear memory updating, amygdala, periaqueductal gray (PAG) and hippocampus are found to play an important part in PE detection and computation. The prefrontal cortex (PFC) and its subregions play a crucial role in the process of PE-initiated memory reconsolidation. Furthermore, some essential neurotransmitters in the nervous system are involved in this process, notably dopamine and glutamate. In the future, quantitative investigations based on statistical calculation models of

PE need to be conducted to explore the interactions between PE and other boundary conditions on memory reconsolidation. The role of different types of salience in memory reconsolidation is also worth investigating. In addition, individual difference in PE's role in updating fear memories must be taken into account to facilitate clinical translations. In both basic research and therapeutic intervention attempts, we feel that multidisciplinary techniques and procedures are essential for elucidating the processes underlying the involvement of PE in fear memory reconsolidation and updating.

Key words: prediction error, fear conditioning, memory updating, reconsolidation, reconsolidation interference paradigm